ARI Research Note 2003-10

Virtual Environment Cultural Training for Operational Readiness: VECTOR

William J. Weiland
John Deato
Charles A. Barba
Thomas P. Santarelli
CHI Systems, Inc.

April 2003

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

ZITA M. SIMUTIS
Director

Technical Review by

Michael J. Singer, ARI

NOTICES

DISTRIBUTION: Primary distribution of this Research Note has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army

Research Institute for the Behavioral and Social Sciences, Attn: TAPC-ARI-PO, 5001 Eisenhower Ave., Alexandria, VA 22333-5600.

FINAL DISPOSITION: This Research Note may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Note are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) April 2003	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) March 2002 – December 2002
4. TITLE AND SUBTITLE Virtual Environment Cultural Training for Operational Readiness: VECTOR		5a. CONTRACT OR GRANT NUMBER DASW01-02-C-0013
		5b. PROGRAM ELEMENT NUMBER 0605502A
6. AUTHOR(S) William J. Weiland, John Deaton, Charles A. Barba, and Thomas P. Santarelli		5c. PROJECT NUMBER 2O665502M770
		5d. TASK NUMBER 6901
		5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) CHI Systems, Inc. 716 N. Bethlehem Pike Suite 300 Lower Gwynedd, PA 19002-2650		8. PERFORMING ORGANIZATION REPORT NUMBER 021223.02008
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI
		11. MONITOR REPORT NUMBER Research Note 2003-10
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution authorized to U.S. Government agencies and their contractors to protect proprietary information. 13 January 2003. Other requests for this document shall be referred to the U.S. Army Research Institute.		
13. SUPPLEMENTARY NOTES Technical Monitors: Bruce W. Knerr and Michael J. Singer		

14. ABSTRACT *(Maximum 200 words)*:

This report describes the effort and results of a Phase I SBIR research program known as Virtual Environment Cultural Training for Operational Readiness (VECTOR). The purpose of VECTOR is to enable the use of virtual environments for training cross-cultural skills, such as language and customs, by providing synthetic actors that exhibit correct cultural behaviors. Accomplishing this task draws on technologies for modeling cognition and emotion, and automated speech recognition and synthesis. A key thrust of the Phase I effort was to develop approaches and techniques to represent cultural factors and embed these within existing cognitive modeling technologies that could drive synthetic actors and associated avatars in a virtual environment

15. SUBJECT TERMS Virtual environments, cultural familiarization, cognitive modeling, training, speech recognition, speech synthesis

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited		Dr. Bruce W. Knerr (407) 384-3987

ARI Research Note 2003-10

Virtual Environment Cultural Training for Operational Readiness: VECTOR

William J. Weiland
John Deaton
Charles A. Barba
Thomas P. Santarelli
CHI Systems, Inc.

Simulator Systems Research Unit, Orlando FL
Stephen L. Goldberg, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

April 2003

Army Project Number **Small Business Innovative Research**

2O665502M770

Distribution authorized to U.S. Government agencies and their contractors to protect proprietary Information. 13 January 2003. Other requests for this document shall be referred to the U.S. Army Research Institute.

FOREWORD

During the last decade, the U. S. Army has been required to conduct an increasing number of missions. Many of these missions represent significant departures from the Army's primary warfighting mission. They include missions conducted to enforce peace agreements, and to provide humanitarian assistance and disaster relief. These missions, often carried out as part of a multi-national force, require leaders and soldiers to possess skills in addition to those required for success in combat. They often require junior leaders and soldiers to interact and communicate personally and effectively with people whose cultures, languages, lifestyle, and beliefs are very different from those found in the United States.

Virtual Environment (VE) technologies have the potential to provide the Army with a training capability to meet these new demands. Immersive virtual simulations have been shown to be effective for training a variety of tactical skills. However, tactical training generally does not require the level of fidelity in representation of interpersonal interaction that is required for cultural familiarization.

In response to this need, the U.S.Army Research Institute for the Behvavioral and Social Sciences, Simulator Systems Research Unit (ARI-SSRU) prepared a topic for inclusion in the DoD Small Business Innovation Research Program entitled "Virtual Simulation Tools for Cultural Familiarization." This report describes the results of the first phase of a research effort conducted under that topic by CHI Systems, Inc. to develop and evaluate tools and methods which make virtual simulations a more effective means of training military personnel to work and interact effectively with people of other cultures. The results form the basis for a subsequent Small Business Innovation Research Phase II effort.

Kathleen A. Quinkert
Acting Technical Director

EXECUTIVE SUMMARY

Research Requirement:

This report describes a Phase I Small Business Innovation Research (SBIR) program known as Virtual Environment Cultural Training for Operational Readiness (VECTOR). The purpose of VECTOR is to enable the use of virtual environments (VE) for training cross-cultural skills, such as language and customs, by providing synthetic actors that exhibit correct cultural behaviors. Soldiers often perform missions requiring them to teach, negotiate, guide, and lead people from different cultures. To be successful in such missions, soldiers must first of all possess an understanding of these cultures, and the cultural communication skills necessary to work effectively with their counterparts within the host nation, and other coalition forces, all of whom have cultures different from our own. The key requirement of the Phase I effort was to develop approaches and techniques to represent cultural factors and embed these within existing cognitive modeling technologies that could drive synthetic actors and associated avatars in a VE.

Procedure:

The VECTOR concept was developed to meet this requirement for cultural familiarization training through the application of highly experiential, scenario-based training in VE that could be used to develop specific skills for interacting with members of a culture of interest. Such training would directly support asymmetric warfare requirements in that all personnel assigned to a particular region/country/operation would be better prepared to deal with decisions linked to an unfamiliar social context.

A number of key technology areas and types of expertise were considered. These technologies included: VE/3D graphics development, speech recognition and synthesis, natural language processing, cognitive modeling, intelligent tutoring/training systems, and advanced interaction technologies (including body/head/eye tracking and haptic interaction). The concept was to employ such technologies in the design and development of synthetic actors to represent the members of the other culture.

Findings:

The overall principle of training envisioned in VECTOR combines notions of constructivist, case-based learning with guided practice. In this approach, a trainee is placed in a scenario within a simulated environment. In the early phases of learning, the trainee is provided with behavioral modeling (i.e., behaviors are demonstrated and explained) and significant ongoing feedback and coaching. As instruction progresses, the amount of continuous feedback is reduced, until the trainee progresses through the entire scenario without intervention, and is only provided with an after-action review (AAR) at the end. The degree of intervention is based on the trainee's evolving skill level and amount of exposure to the material.

Wise's cultural template (Wise et al., 1997) provided the basis for development of a cultural model in VECTOR. Effectively, it posited the existence of a broad set of common cultural variables, validated with respect to their practical impact on military missions, that could be identified for a range of cultures. In addition to this basic framework, the study identified useful cultural behavioral rules that can be utilized in constructing training scenarios and interaction scripts.

One of the principal determinants of the successful application of VECTOR as a training technology is the tractability of quickly generating new cultural models, when required, and generating training scenarios and interaction scripts to carry out training. An approach we identified to facilitate such rapid generation of new training material is to provide for a layered modeling approach that provides for extensive model reuse.

Training cross-cultural interaction within a VE imposes certain requirements on the interaction techniques used within the VE. A great deal of the interaction, and the opportunities for creating misunderstandings and conflicts, depend on relatively subtle cues, such as physical proximity and posture, gaze direction, use of touch, and tone of voice. Unfortunately, detection of many of these cues is problematic given the state of VE technology, and depending on the design of a given VE, may not be feasible in any case. To the extent possible, we are relying on off-the-shelf interaction technology; where necessary we plan to adapt or improve existing technology where required by the interaction requirements of the initial target training environment.

Utilization of Findings:

Concepts and techniques identified during Phase I will be further developed, demonstrates, and evaluated during Phase II. Technological issues to be addressed include integrating the cultural model into a VE along with speech processing components. As much as possible, technology leveraging from related projects will facilitate the design, development, and modeling effort.

CONTENTS PAGE

LIST OF FIGURES

VIRTUAL ENVIRONMENT CULTURAL TRAINING FOR OPERATIONAL EFFECTIVENESS: VECTOR

This report describes the efforts and results of a Phase I SBIR research program known as Virtual Environment Cultural Training for Operational Readiness (VECTOR). The purpose of VECTOR is to enable the use of virtual environments for training cross-cultural skills, such as language and customs, by providing synthetic actors that exhibit correct cultural behaviors. Accomplishing this task draws on technologies for modeling cognition and emotion, and automated speech recognition and synthesis. A key thrust of the Phase I effort was to develop approaches and techniques to represent cultural factors and embed these within existing cognitive modeling technologies that could drive synthetic actors and associated avatars in a virtual environment (VE).

This report begins with a description of the VECTOR concept, describes the cultural analysis and a training scenario, then discusses our approach to cognitive modeling and extensions to represent cultural factors. It concludes with some notes on software architecture and implementation.

Concept

The number of military activities classified as "operations other than war" (or OOTW) is on the increase. Often called asymmetric warfare, these activities may include such things as counter-insurgency campaigns, hostage rescue operations, low intensity conflicts, military operations in urban terrain (MOUT), and peacekeeping operations. Such activities require a vastly different set of tactics, equipment, training and skills than conventional military engagements of the past. Future conflicts may not involve commitments of massive numbers of troops to fixed battle zones, but will likely involve combating small units of fanatical terrorists wielding Weapons of Mass Destruction (WMD) and other sophisticated tactics and technologies. Moreover, these missions will require leaders and soldiers to possess a different set of skills from what was required for success in traditional combat situations. They will most likely require leaders at all levels of command to interact and communicate personally and effectively with people whose cultures, languages, lifestyles, and beliefs are very different from those found in the U.S.

Recent U.S. military experience in Bosnia resulted in a series of lessons that should inform U.S. military policy in future situations of this kind. One of the more important lessons to come out of the Bosnian experience was the need to refocus training and develop senior military leaders for participation on peace operations (Olsen & Davis, 1999). U.S. military officers were confronted with new problems in Bosnia that included non-traditional challenges such as negotiating with factional leaders and local government officials, managing civil-military relations, and securing a safe environment for implementation of the Dayton Peace Accords. In response to his experiences in Bosnia, Army Chief of Staff Gen. Erik Shinseki was quoted as saying that he had to confront a "cultural bias"…"there wasn't a clear doctrine for stability operations. We are developing it, using the Bosnia experience, to define a doctrine for large stability operations. But it is this absence of a doctrine for a doctrine-based institution that you walk into in this environment. There you are in a kind of roll-your-own situation" (quoted in Olsen & Davis, 1999). According to Olsen & Davis, the peace operations in Kosovo required skills such as patience, the confidence to delegate authority and take risks, and the ability to communicate with people outside the military, including representatives of nongovernmental and international organizations and the media. The authors' conclusion emphasized a need to develop a set of general skills that enhances all levels of officer education, to include cultural awareness and interpersonal skills. Preparation for leadership roles in future peacekeeping missions like the Bosnian operation will present a critical need in the upcoming decade. If these conclusions have been drawn for the officer corps, it is a safe assumption that the level of cultural awareness and interpersonal skill ability of senior enlisted personnel is also in need of training.

Soldiers often perform missions requiring them to teach, negotiate, guide, and lead people from different cultures. To be successful in such missions, soldiers must first of all possess an understanding of these cultures, and the cultural communication skills necessary to work effectively with their counterparts within the host nation, and other coalition forces, all of which probably have cultures different from our own. While the U.S. should have tremendous advantages from its advanced information and battlefield management systems, we are unprepared from a personnel readiness standpoint to deal with people from different cultures. This state-of-affairs makes our soldiers prone to making ill-advised decisions that potentially undermine our ability to successfully complete missions that involve interacting with members of a different culture.

The VECTOR concept was developed to meet these needs by providing a new technology for training in cultural familiarization through the application of highly experiential, scenario-based training in VE, that could be used to develop specific skills for interacting with members of a culture of interest. Such training would directly support asymmetric warfare requirements in that all personnel assigned to a particular region/country/operation will be better prepared to deal with decisions linked to an unfamiliar social context.

A number of key technology areas and types of expertise were considered in developing such a training system. These technologies included:

a. virtual environment/3D graphics development,
b. speech recognition and synthesis,
c. natural language processing,
d. cognitive modeling,

e. intelligent tutoring/training systems, and

f. advanced interaction technologies (including body/head/eye tracking, and haptic interaction).

The concept was to employ such technologies in the design and development of synthetic actors to represent the members of the other culture. VECTOR was initially formulated and built to deliver instruction in critical decision-making skills as well as opportunities to practice these skills with synthetic agents and instructors in a vignette-oriented virtual simulation. The development of training technology such as this comes with several important technological challenges. In order to meet those challenges, it was decided during our Phase I work to delegate support for VE, interaction devices, and avatars to third party implementations to the extent possible. That way, future work could focus primarily on additional cultural and cognitive model support, interaction scripting for training content, and support for speech interaction. Using this approach, resources can be devoted to overall synthetic actor model development that will enable significant reuse of cultural models in adaptation of VECTOR to new cultures and situations.

The Phase I effort made significant progress in both the identification of the training domain and practice scenario. A cultural model framework was initiated in Phase I to demonstrate model execution. Technological issues to be addressed include integrating the cultural model into a VE along with speech processing components. As much as possible, technology leveraging from related projects will facilitate design, development, and modeling efforts.

Instructional Design

As the previous sections described, the problem is training soldiers in cultural understanding as it relates to the performance of their duties in foreign locales. The approach is based on immersive VE, featuring, to the extent possible, naturalistic interaction (natural language speech and gesture), and involving scenario-based instruction. While Special Forces-like missions and personnel have been selected as the initial target audience, it is anticipated that the ultimate benefits to training conventional forces will be compelling, and potentially far greater. Special Forces training normally includes intensive cultural training, involving extended periods living and interacting with indigenous peoples in selected regions of the world. While VECTOR offers the possibility of effective, intensive instruction in advance of such real-world experience, it is unlikely that any computer-based training system can offer the depth of understanding obtained from living within a culture. On the other hand, conventional forces will increasingly be required to undertake missions in which cross-cultural interactions play a significant role in determining success or failure. These conventional forces will not have the benefit of intensive real-world experience, and may in fact need to be deployed with limited forewarning. A brief, intensive, effective, and rapidly-adapted training program in cultural interaction is what VECTOR is intended to deliver, which should benefit conventional forces subject to time constraints and training needs on a large scale.

Cognitive Apprenticeship

Educational research has shown that active student involvement produces better results, particularly when the goal of the instruction is acquiring some practical skill (e.g., as opposed to learning abstract concepts). A common way of providing this active participation is to allow students to practice the skill in real or simulated settings. The heavy and high-payoff investment that the aviation industry has made in flight simulators provides highly successful examples of this principle in action.

However, practice by itself is also insufficient, even when used to reinforce didactic instruction. Teaching should be tailored to individuals, an idea that has a long history in education. A great deal of data shows that carefully individualized instruction is superior to conventional group instruction, and this is particularly the case in active or skilled processes. In an article dealing with applied skill training, Gott (1989) discusses the need for training methods that are a combination of modern understanding of skill acquisition and traditional apprenticeship training techniques, such as mentoring and coaching. “Cognitive apprenticeship” is a model of instruction developed by Collins (1987), who adapted the traditional view of apprenticeship in learning physical skills in tangible activities to learning less tangible cognitive skills. Cognitive apprenticeship instruction involves taking trainees through successive approximations of mature practice as they learn how to perform job tasks. The important aspects of cognitive apprenticeship training are:

- realistic practice environments and problems, in which students can execute realistic tasks in a realistic environment and see how knowledge (acquired through didactic instruction) can be practically put to use;
- support from the tutor, including real-time comments on student performance, providing hints, reminders, and explanations;
- a gradual fading out of this support as skill builds; and
- careful sequencing of problems and challenges that are appropriate for the growing skill level of the learner.

In his discussion, Gott (1989) concludes that technical skill development is best achieved by “guided experience in instructional environments that provide progressive, explanation-based, and otherwise generally supported practice in the mechanics of solving problems.” This overall method, which combines didactic instruction and individually-mentored active practice, can be simply termed ‘guided practice’.

Training cultural knowledge requires more than merely inculcating a set of behavioral rules in a trainee. To some extent, the trainee needs to internalize a mindset; a specific set of cultural encounters in a training scenario will never serve to cover all possible cases that a soldier may encounter in real life, but needs to be designed to generalize into a broader understanding. In this sense, the training approach selected for VECTOR needs to be constructivist. There is significant academic work that establishes the validity of this kind of approach; for example, Schank's work in Case-Based Reasoning has been extended to Case-Based Learning ((CBL); Riesbeck, 1996), where learning is conceived of as a process of compiling cases (specific examples of actions and results) and constructing an index of cases to apply to novel situations (being able to generalize appropriately from cases). CBL implies that trainees are exposed to a broad base of experiences, in compelling and memorable environments. In addition, being able to index those experiences requires that motivation (underlying goals and plans) be present, that both failure and success be experienced, and that reasons for failure be understood.

In applying the CBL approach to VECTOR, a VE provides the necessary vivid and realistic experiences. In addition, a concept of mission (a mission briefing or orders by a virtual commanding officer that convey higher level commander's intent as framing goals, for example), and strongly branching scenarios that permit failure (bad decisions) to be followed flexibly to their conclusion, can provide the basis for effective CBL. In terms of instructional design, the overall principle of training envisioned in VECTOR combines these principles of constructivist, case-based learning with guided practice. In this approach, a trainee is placed in a scenario within a simulated environment. In the early phases of learning, the trainee is provided with behavioral modeling (i.e., behaviors are demonstrated and explained) and significant ongoing feedback and coaching. As instruction progresses, the amount of continuous feedback is reduced, until the trainee progresses through the entire scenario without intervention, and is provided with an after-action review (AAR) at the end. The degree of intervention is based on the trainee's evolving skill level and amount of exposure to the material.

Scenario-based Training

The notion of scenario in VECTOR bears some explanation. Rather than being conceived of as a rigid and potentially brittle sequence of actions that take place in the environment on the basis of fixed times, the scenario represents a collection of players in the environment that have predefined roles, tasks, and motivations. Some external events may happen at fixed times (e.g., an air strike on a building); other events may happen on the basis of proximity or other contingencies (e.g., walking down the wrong street may result in being attacked). Essentially, however, the scenario plays out solely on the basis of the trainee's actions in the environment. In this respect, it has much in common with current game technology.

Control over the interventions, and the content of those interventions, must be provided by a tutoring agent. The tutor can be embodied as a soldier who accompanies the trainee through a scenario; alternatively, the tutor can "channel" through a number of avatars (such as a commanding officer for initial mission briefing and AAR, or other soldiers or actors during the mission). In a full-fledged implementation of this concept, the tutor would be monitoring interactions and maintaining a set of Measures of Performance (MOP's) and Measures of Effectiveness (MOE's). MOP's might include behaviors indicative of a particular skill being trained and the percentage of their correct vs. total applications. MOE's might include time to completion of the scenario, achievement of mission parameters (successful achievement of objectives, number of casualties, etc.).

An AAR could take several forms, including tabular display of MOP/MOE statistics, an overall rating, and list of good and bad actions (possibly provided by an avatar). In addition, a narrated replay of key decision points in the mission might be an effective tool for reinforcing learning and indexing.

In order to strike a reasonable balance between instructional generality and timely implementation, we will implement the instructional system and tutoring component in stages. The first stage will provide a mission briefing and scenario-based execution of the mission. Next, basic metrics will be maintained and used to construct a summary AAR. As scoring of the trainee's individual behaviors is implemented, the notion of explanation can be implemented (basically, when things went badly, the cognitive state of the actors involved can be analyzed to determine why particular actions were bad, e.g., "when the trainee said X, the actor was offended," etc.). iGEN has an explanation facility that may be used to implement this type of functionality. In turn, we would use this to provide some types of simple feedback during the mission. At this point, the VECTOR system would have several operating modes, ranging from "model behaviors and provide correction," to "provide limited feedback," to "provide only AAR."

Future evolutions of the system might take this further by developing a full-fledged tutor that could use this kind of information to manage the scenario more carefully, ensuring that trainees do not depart too far from the scenario, or adaptively modifying the level of feedback based on performance and trainee history.

Cultural Analysis

In order to be effective in OOTW**,** MOUT, and peacekeeping operations, soldiers must possess an understanding of the culture and the appropriate cultural communication skills necessary to work effectively both with friendly forces and any indigenous population. This type of interaction has usually been relegated to Special Forces personnel who have a demonstrated ability to succeed in these situations. The potential for U.S. forces to be involved in asymmetric warfare continues to increase. It is inevitable that conventional forces will be required to participate in missions usually reserved for Special Forces. This will create a dilemma. How can conventional forces quickly train to deal with the challenges created by asking soldiers to be sensitive to a different and unfamiliar culture as part of their mission objectives? Although this question cannot be definitively answered, the VECTOR concept can provide a significant benefit to augment the standard cultural familiarization methods (i.e., "do's and don'ts" booklets).

Culture is defined as the set of traditional behaviors that are passed down through generations within a specific segment of the human race, who share a racial or ethnic identity, a geographic location, and a specified time period (Wise et al., 1997). The information about a particular culture is usually voluminous; however it is rarely expressed in terms of a list of acceptable or unacceptable behaviors. An anthropologist can devote a significant amount of time to studying the available literature to develop guidance for appropriate cross-cultural behaviors. The value of cultural familiarization is obvious. Soldiers on OOTW and peacekeeping missions are required to teach, negotiate, lead, and maintain the order of people from a different culture. A soldier with cultural training can be more effective on his or her mission while minimizing the possibility that his or her actions may inflame or otherwise weaken a fragile situation.

In order to develop an effective demonstration of the value of cultural training and the usefulness of our VECTOR technology it was necessary to target a specific culture and operational mission. After an extensive literature review and discussion with our sponsor it was decided that our Phase I effort would be directed at peacekeeping operations within the cultural context of the Kurdish population of Northern Iraq. This choice was based on the likelihood of future military operations in this region and the realization that this sub-culture is substantially different from the overall parent Arab culture. U.S. military experience with the Kurdish culture is limited. At the beginning of Operation Provide Comfort, Special Forces soldiers were not area, culturally, or language oriented, since Northern Iraq was not a normal area of operation (Nash, 1997). Operation Provide Comfort was largely successful due to the fact that Special Forces personnel were highly trained in terms of interpersonal and cultural skills and were able to quickly assimilate and apply knowledge about the Kurdish culture. The outcome may have been very different if large numbers of conventional forces were tasked with a similar mission.

A major product of the Phase 1 research was the development of a training scenario, suitable for implementation in a VE, which demonstrates the application of cultural knowledge to enhance mission success. The scenario was developed and reviewed by both a retired Ranger officer and a Special Forces officer to insure operational realism and relevance.

Cultural information on the Kurds was collected from various sources. Our primary source of cultural information was a product of an Army Research Institute program to identify and prioritize Special Forces cultural communication skills and training needs (see Wise et al., 1997). That research effort developed a systematic methodology to identify and prioritize cultural communications skills and behaviors. Their methodology used a generic cultural template that included criteria to direct the collection and cataloging of information on any specific culture of interest. Such criteria included:

a. Cultural Core/Psychological Profile -- which encompasses cultural values, relationships, and motivations;

b. Shared History -- which involves traditions, social structures, manners, linguistics, the arts, science, culinary traditions, leisure, sports, hygiene, and minority groups; and

c. National Considerations -- which involves geography, politics, legal system, military, economic structure, finance, commerce, housing, health, communications, transportation, and education.

The methodology was then applied to create lists of behaviors specific to Asian, Arab, Russian, and Cuban cultures. As part of developing this cultural template, the authors (and a panel of experts born and raised in the specific culture), prioritized this list of behaviors with respect to a specific military mission or operation. Highly skilled Special Forces personnel participated in this latter step. Each item in the template was ranked by its military importance with respect to a given specific mission. As a result of this analysis, the prioritized listing of cultural communication skills for the Arab culture contained over 400 behaviors. Although Kurdish culture differs significantly from Arab culture, both Kurds and Arabs share a large set of common secular and religious customs, norms and behaviors.

A key feature in the research of Wise was the development of a generic cultural communications skills template. The generic cultural template is a taxonomy or representation method, which can be used to organize information about any culture. The cultural template provides us with a systematic approach for the logical representation of communication behaviors and their effect in a specific culture. The research done by Wise can be considered the definitive attempt to categorize and prioritize the cultural communication skills and their importance in training for military missions.

Additional cultural information pertaining to the Kurdish culture was collected from various U.S. Department of State, Bureau of Population, Refugees, and Migration sponsored websites as well as several books written on Kurdish culture (see Kreyenbroek & Allison, 1996; Meho & Maglaughlin, 2001). The remainder of this section provides a detailed description of the training scenario that was developed. This scenario has been developed by CHI Systems, but was based extensively on an extrapolation of the seminal work performed by Wise et al., (1997) as summarized above. That is, important communication skills for the Arab culture reported by Wise were woven into the training scenario that will be presented. It is important to note that these training examples were based upon solid experimental findings, and not randomly selected on face value only. The scenario uses Special Forces personnel on an intelligence collection mission to demonstrate the usefulness for cultural knowledge and the potential consequences. Our intent is that the VECTOR concept and scenarios will be initially validated using Special Forces missions before it is applied to conventional forces missions. We feel it is appropriate to utilize special forces expertise in the development of training technologies prior to their introduction to conventional forces.

Phase 1 Scenario Background

U.S. ground forces have entered Northern Iraq as part of a multi-national force. Active combat operations are continuing in the areas of Baghdad and in Southern Iraq, however, large-scale organized opposition has ceased. The areas north of the 36th parallel are considered secure. After directly supporting the U.S. advance, Special Forces A Teams have been relieved from the line and reassigned to intelligence collection operations in occupied portions of Iraq. One particular A Team has been sent to the city of Mosul and its surrounding villages.

Scenario Setting

The simulation takes place in the mid-sized village of Dahuk, which is on the outskirts of the city of Mosul in Northern Iraq. Dahuk consists of approximately 30 small family dwellings and a few small buildings. The village contains a Mosque, a small market place, and a UN humanitarian clinic.

The village is inhabited exclusively by Kurds but contains two factions or tribes: the PKU (Patriotic Union of Kurdistan) and the KDP (Kurdish Democratic Party). Affiliation to these tribes is based on genetic as well as political heritage. The PKU and KDP are distinct tribes. They may cooperate with each other but frequently have disagreements based on ideology and tribal allegiances. Both the KDP and PKU have a hatred for non-Kurdish Iraqis. The KDP and PKU are easily distinguished by their clothing. The KDP traditionally wear green while the PKU wear yellow.

The city of Mosul is also predominantly Kurdish but has a substantial non-Kurdish minority. This minority consists of Iraqis sympathetic to the Saddam Hussein regime. The Iraqis as well as both Kurdish factions are Sunni Muslims. Although the Kurds and Iraqis share a common religion they have little else in common and are frequent adversaries. After the recent deployment of U.S. forces, militant members of both the PKU and KDP are suspected of reprisals against Iraqis in Mosul.

The mission of the A Team in this area is intelligence collection and general peacekeeping. Although the U.S. forces are more sympathetic toward the Kurdish population, they are charged with protecting both Kurds and non-Kurds alike. Acts of terrorism and intimidation against non-Kurdish Iraqis have continued in Mosul. U.S. Forces have been tasked with reducing this Kurdish aggression. Radical elements of the PKU are largely responsible for the violence. The mission of the A Team is to maintain the peace and collect intelligence on the radical elements of the PKU. The radical elements of the PKU have been given considerable support from the PKU citizens of Dahuk. Until this point no Kurdish aggression has been directed against U.S. forces.

The A Team has sent small teams of soldiers to collect intelligence in the area of Dahuk. Each team consists of three men, all of whom speak Kurdish. Each team is led by a warrant or commissioned officer. The teams are interested in obtaining information regarding the whereabouts of the radical elements of the PKU. The teams can expect little cooperation from PKU affiliated civilians but may receive some cooperation from KDP civilians. While collecting intelligence, the teams are instructed to maintain good relations with the citizens of Dahuk and not to offend or otherwise create a hostile situation.

The Training Scenario

The simulation opens with the three-man team approaching the first house on the edge of the Dahuk village. The team consists of CW3 Ford (the team leader), SFC Miller, and SSG Jones. It is early afternoon. No one is present in front of the house and the door is closed. The soldiers can hear people off in the distance but no one is initially visible.

CW3 Ford approaches the door at the first house and he knocks. A woman opens the door. CW3 Ford asks if her husband is home (this is the correct action, conversations with women should be avoided). The man of the house comes to the door. CW3 Ford notices he is wearing a yellow vest over his shirt. The yellow vest is a sign of the PKU tribe and CW3 Ford realizes that he will get little information here. CW3 Ford greets the man at the door and offers his hand for a handshake (a handshake is appropriate between and within the sexes in Kurdish culture - however, it is inappropriate for a man to shake the hand of an elderly Kurdish woman). CW3 Ford introduces himself and shakes hands. CW3 Ford asks the man if he has seen any men with guns in the town. The Kurdish man politely says no. CW3 Ford thanks the man and the team moves on to the next house.

CW3 Ford and the team walk over to the next house and they knock on the door. An elderly man answers the door. The elderly man is also wearing yellow and CW3 Ford notices a pistol tucked in the man's belt. CW3 Ford is concerned. He asks the man for permission to enter the house, but the elderly man refuses. The elderly man becomes agitated. CW3 Ford decides to back off. It is unlikely that the elderly man is a PKU militant and CW3 Ford recalls from his rules of engagement "Possession of a weapon by an individual does not, by itself, constitute a hostile intent" (CALL, 1995). The team moves on down the street.

As the team approaches the next house the door opens and two men walk out. The men shake hands and kiss on the cheeks. One man departs. (It is socially acceptable for men to kiss and show affection between family members in Kurdish culture). The younger SSG Jones is tempted to make a comment regarding the kiss as a sign of homosexuality. (This would be a disastrous mistake; “never imply that someone is homosexual” is ranked as number 4 on the listing of Arab cultural behaviors [see Wise et al., 1997]). SSG Jones says nothing. CW3 Ford approaches the man who is wearing green (a sign of the KDP tribe). CW3 Ford feels good about his chances for getting some information and he extends his hand for a handshake. CW3 Ford introduces himself and shakes hands. CW3 Ford begins to describe his concerns about PKU militants in the town to the man. While CW3 Ford is talking he is moving his head and scanning the surrounding yards for activity (a typical soldier behavior). CW3 Ford is not making eye contact with the man as he speaks. In Kurdish culture it is very important to make and maintain eye contact during a conversation. The man becomes disinterested in what CW3 Ford has to say. The man apologizes, says he can’t help, and returns to his house. CW3 Ford realizes that if he had been more attentive he may have gotten some useful information.

The team moves on to the fourth house. CW3 Ford approaches the door and knocks. A man wearing a green vest answers the door; CW3 Ford assumes he is a member of the KDP. CW3 Ford offers his hand for a handshake, the man accepts. CW3 Ford begins to explain to the man why he and his team are in the village. The man is interested in what he has to say and he invites CW3 Ford and SFC Miller into his home, SSG Jones remains outside. As soon as the soldiers enter the home, an old woman immediately begins to become very agitated and yells at the soldiers. Hoping to defuse this situation, CW3 Ford offers his hand to the woman. This is a mistake. Although it is acceptable to shake hands with Kurd women, usually forbidden in other Arab sub cultures, it is inappropriate to ever touch an elderly Kurdish woman. Noticing CW3 Ford’s intent, the Kurdish man steps between CW3 Ford and the old woman. The Kurdish man manages to get the woman to leave the room. The two soldiers and the man begin to talk more. CW3 Ford explains that he is seeking information on the whereabouts of radical PKU members. The man says that he distrusts the PKU and offers useful information to CW3 Ford. He tells CW3 Ford that he believes that a group of armed PKU members are in house 5 right now. CW3 Ford thanks the man for the information and he and SFC Miller leave the house. Once outside, CW3 Ford informs SSG Jones of what he learned.

The team crosses the street and approach the fifth house cautiously. CW3 Ford orders his two men to stay back as he approaches the door. CW3 Ford knocks loudly on the door. No one answers the door but CW3 Ford continues to knock. Suddenly, SSG Jones yells out “four armed men running from the back of the house!” CW3 Ford immediately yells, “Take Cover!” The men fall into a prone position, CW3 Ford runs back to where his men are located. CW3 Ford orders SFC Miller to radio back to company headquarters that they suspect armed PKU presence in Dahuk. The team is advised to take no aggressive action unless fired upon, to proceed on their intelligence collection mission, and are informed that MP units have been dispatched and are in route to Dahuk.

CW3 Ford and his men proceed to the back of the house. They notice lots of footprints leading from the back door to the mosque next to the fifth house. The team begins to approach the front of the mosque. CW3 Ford realizes that he cannot enter a mosque without permission (ranked 7th on the list for cultural behaviors [see Wise et al., 1997]) and he begins to knock on the mosque door. No one answers the door. Believing that the armed PKU members are in the mosque, he begins to knock louder and he yells "is anyone in there!" CW3 Ford should not have yelled outside a mosque (ranked 16th on the list for cultural behaviors [see Wise et al., 1997]), this is both a cultural and religious violation. At this point a crowd of townspeople from the nearby marketplace begin to assemble around the team. The crowd is not at all happy with the team's presence in Dahuk. CW3 Ford realizes this and orders his team to withdraw. The team withdraws and waits on the edge of town for the MPs to arrive.

Cultural Modeling in VECTOR

Having discussed how cultural differences can affect interactions with indigenous populations, we can next consider how these can be represented, and how they can be dealt with computationally in a training system. The focus of the VECTOR effort is to develop training approaches for culture, rather than purely modeling culture, so we look at this from the perspective of engineering a solution to the cultural training problem. One of the potential problems with attempting to model culture at a deep level and having that induce culture-specific behaviors in the training context is that it can be difficult to validate these models, and difficult to predict how and when these behaviors will emerge during training. Another thing to consider is that dialog with individuals is the primary mechanism by which cultural factors come into play. This leads us to consider the option of bringing a large part of the cultural model into the dialog itself. Effectively, an interaction script is constructed based on anticipated dialog that could occur in the training scenario. Each utterance received from the trainee is associated with a set of utterances that could be generated in response by a synthetic actor in the VE. The synthetic actor's cognitive and emotional state, based on prior interactions and events in the environment, affect the choice of responses from the script. Subsequent sections will provide greater detail on the processing of the interaction script, particularly in light of its limited ability to represent the broad range of possible naturalistic interactions.

Cultural Model Framework

Wise's cultural template (Wise et al., 1997) provided the basis for development of a cultural model in VECTOR. Effectively, it posited the existence of a broad set of common cultural variables that could be identified for a range of cultures. The set of cultural variables was validated with respect to their practical impact on military missions. Thus, the result of this study provides a useful underpinning to the VECTOR work. In addition to this basic framework, the study identified useful cultural behavioral rules that can be utilized in constructing training scenarios and interaction scripts. In this section, the developed cultural model is described.

One of the principal determinants of successful application of VECTOR as a training technology is the tractability of quickly generating new cultural models, when required, and generating training scenarios and interaction scripts to carry out training. An identified approach to facilitating such rapid generation of new training material is to provide for a layered modeling approach, that provides for extensive model reuse. The first exemplar of a VECTOR cultural model will be implemented in future efforts, and should permit ready adaptation to other cultures. The layered model comprises four levels:

- cultural template, that specifies generic, parameterizable cultural attributes that define broad, largely universal characteristics of a culture;
- cultural specialization, that provides for definition of idiosyncratic cultural norms and behaviors, as well as recognition of gesture and dialog cues;
- individual specialization, that further provides for definition of individuals within a training scenario and provides such information as gender, name, specific goals and intentions, and that can establish offsets from the parameters provided at the higher levels of the model; and
- individual memory, that provides recall of specific interactions that took place within the context of a scenario.

As mentioned above, the dialog itself represents a significant portion of the encoded cultural knowledge. Language is one of the main cultural differentiators, but in addition to the obvious differences in language, the dialog (interaction script) can encode culturally-determined linguistic responses to gestures and other extra-linguistic cues. The intent is that the responses be conditional on the basis of variables such as animosity and level of trust, that can both reflect cultural norms (e.g., characteristic distrust of strangers) and be the result of disregard of correct protocols (e.g., dislike based on offensive behaviors). A final characteristic of these dialogs is that they be threaded, to permit a back-and-forth exchange to be encoded (for AAR).

As mentioned previously, the intent of the dialog is that a simple, readily-adapted encoding of both cultural information and the training scenario be available. Effectively, the dialog that is constructed by the training developer represents a sample interaction, with some possible branching built in. The ability of the combined dialog system and synthetic actor/cultural model to perform loose matching against the trainees utterances, and to select responses from the dialog based on the history of the interaction means that robust, yet not excessively constrained, training interaction is possible.

The individual synthetic actors are controlled by a hybrid model comprised of three components:

- a perceptual model, consisting of ad hoc software recognizers for gestures and physical actions that are of importance in particular cultural contexts, or within particular implementations of virtual environments;
- an emotional model, based on existing theoretical models that posit "basic" emotions, extended with a set of emotional parameters that are chosen for pragmatic reasons within the VECTOR context;
- a cognitive model, based on the COGNET framework and implemented within the iGEN toolset (described in greater detail in the next section), that handles decision making and memory for the synthetic actor.

The perceptual model is a set of monitors of VE interaction device data (such as position trackers) that detect and signal physical actions. Examples of this might be a special-purpose detector that monitors the position of hands (for handshakes, waves, or gestures such as the improper use of the left hand in some cultures); or a gaze follower that determines what or who is being examined or addressed (assuming at least head tracking). The purpose of this component is to reduce the volume of data being sent to the cognitive/emotional models and to enable treating them at a higher level of abstraction (i.e., detected a specific gesture, rather than hand at coordinate (x, y, z)). This component may require updating for extension to other cultures, because of the wide range of culture-specific gestures and customs.

The emotional model plays a significant role in the overall reactions and behaviors of the synthetic actors toward the trainee, and forms an important part of cultural training. By ultimately providing the ability to drive facial expressions on the avatars, the emotional model also offers the prospect of more compelling interactions with the synthetic actors. The underlying concept is that we establish a set of variables that represent levels of each of six "basic" emotions. Commonly this set includes happiness, sadness, anger, fear, disgust, and surprise (see Velasquez, 1997). For our purposes, we can choose to augment this set with other emotional variables (such as trust). Such additional variables can be chosen on the basis of their benefit to the realism of training, regardless of their actual theoretical validity. These emotional variables are controlled by a set of (culture-specific) mappings that take output from the gesture detectors, dialog manager, and from events in the environment and apply corresponding adjustments to one or more of these variables. The mechanism can be elaborated to include excitatory and inhibitory effects between emotional variables, as well as saturation levels and a decay curve.

Finally, the cognitive model component is constructed in COGNET/iGEN. This component provides decision making, memory, and drives verbal and physical actions. The cognitive model is responsive to the outputs of the dialog manager and the perceptual model, as well as the states of the emotional variables. Because of its flexibility, iGEN can be used to construct arbitrarily complex behaviors within a synthetic actor, but typically, the cognitive model would be used to decide which responses from the interaction script to emit given the current state of memory, the environment, and emotional factors. So, for example, a lack of trust might cause the cognitive model to decide to lie, or not answer a question posed. A brief description of the capabilities of the iGEN formalism follows.

COGNET/iGEN

COGNET includes a rich language for describing and formalizing knowledge in a form that corresponds to the ways in which it is used by human decision makers (e.g., perception, problem-representation, reasoning, and performance of actions; see Figure 1). COGNET combines several powerful cognitive engineering representations and models, including production rules, pandemonium theory, blackboard representations, and semantic networks (see Zachary et al., 1996). It offers a framework for representing knowledge in a heterogeneous structure. In addition, the iGEN tool offers an implementation of COGNET that provides the capability to use the same structures to efficiently construct intelligent algorithms for non-cognitive-modeling applications (CHI Systems, 2002).

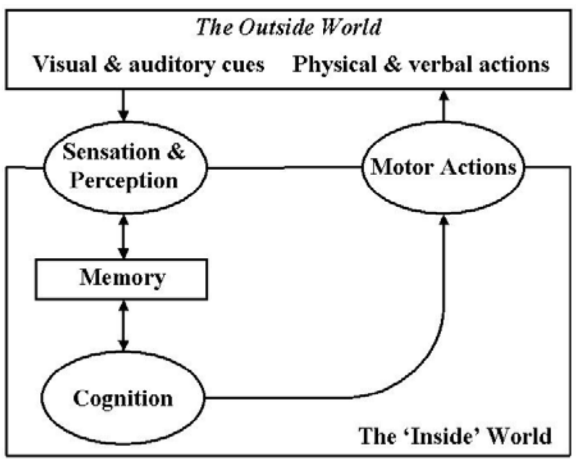

Figure 1. COGNET architecture.

In order to execute a cognitive model, there must be a software simulation of the underlying cognitive architecture and some way to encode the COGNET model in a form accessible to that executable architecture. The software development environment called iGEN performs these functions. iGEN is built around a software implementation of the COGNET information processing architecture. This executable architecture uses domain expertise, encoded in a computational form of the COGNET description language, to process sensory inputs from, and to generate actions toward, some real or simulated problem environment. The inputs/outputs to the external environment are coded in a domain-specific manner using a C++ Application Programmer Interface (API). The API links sensory cues to the COGNET perceptual processes (e.g., perceive speech), and links the COGNET motor processes (e.g., perform action) to system controls that generate environmental effects (e.g., answer question).

In the context of VECTOR, the iGEN cognitive modeling framework is being used to support the intelligent behavior of the synthetic actors in the environment. This includes:

- basic task behavior required by a specific scenario (actions that may be carried out by the synthetic actor);
- interaction memory, recollection of specific data communicated to the actor by the trainee (or by other actors);
- instance data, such as the name of the actor; and
- an emotional model, that defines basic emotional parameters that are used to influence selection of actions and speech.

Speech is not being integrated within iGEN, but rather is handled by an external dialog manager that processes dialog scripts. The concept of the dialog script is not to provide strictly sequenced, "canned" interaction, but rather to avoid the intractable research problem of natural language understanding (particularly across multiple languages). Essentially, the script provides templates for utterances that the synthetic actor can construct, based on variables maintained in the cognitive model, as well as templates for utterances that the synthetic actor can "understand," i.e., match approximately against spoken utterances received from the trainee. The trainee's utterances are matched against these templates, and, where matches occur, parameters (e.g., proper nouns) can be extracted from utterances and stored by the cognitive model. In addition, the overall informational and emotional content of the matched messages is encoded within the script and passed to the model.

iGEN models are basically a type of program code that is executed by a language interpreter (a program). The model "code" defines knowledge, in the form of facts ("hypotheses") that are stored within a blackboard structure (basically, a hierarchy of fact types). The model code also defines tasks, which are step-by-step procedures for accomplishing results, that are themselves invoked when certain trigger conditions (based on the configuration of blackboard hypotheses) occur. Finally, demons and actions provide the perceptual and motor "hooks" of the model to the outside world. An adaptation library ("shell") is linked with the iGEN interpreter ("BATON") to implement these hooks. Each synthetic actor instantiated in a scenario exists as an executing instance of BATON, with a specific model loaded.

Perceptual data, in this case utterances picked up and interpreted by the dialog manager, are passed to the BATON instance(s) that correspond to synthetic actors "listening" to the utterance. These are picked up in the model through the perceptual demons, and are processed and posted to the blackboard, where they can trigger task processing to occur. When processing of a task invokes actions, a flow of information occurs back from the model to the outside world (i.e., through the shell and to the dialog manager, for speech actions, or to the avatar itself for physical actions).

The models to be built for VECTOR will consist of multiple pieces: a generic core that has speech processing and demons/actions for communication with the dialog manager, as well as a cultural and emotional model that can be parameterized (the *cultural template*) when the actor is instantiated to construct basic cultural rules and behaviors. Cultural rules consist of mappings from dialog acts to specific emotional effects (e.g., use of correct address increases trust), as well as specifications of variables such as social distance that the model attempts to maintain.

The basic cultural template consists of a baseline perceptual model, emotional model, and cognitive model with a set of tasks to handle dialog actions and basic actions or interactions (locomotion and greetings, etc.). For a particular culture, the emotional model mapping and activation thresholds in the cognitive model are adjusted. To construct individuals, each individual synthetic actor is specified in a definition file that provides parameter offsets for some of these quantities. In addition, individual characteristics such as name, gender, age, and association with avatar (graphical representation) are entered into the definition and loaded by the associated cognitive model when the actor is instantiated. Instantiation is controlled by a scenario script that defines which actors are "cast" and where they should appear, as well as any ongoing tasks that might be specified for them. Finally, each individual is provided with an interaction script. It may prove desirable to have a basic interaction script that provides default dialog for each actor, but to have an overlaid script that is specialized to each one.

The emotional state (i.e., vector of emotional variables) influences the triggering, priority, and execution of cognitive tasks within the model. A combination of anger and fear, for example, in conjunction with an environmental stimulus (a gun being drawn) might trigger a self-defense task, involving a violent response. Without the presence of anger or fear, a lesser reaction to the appearance of the weapon might be selected by the model (such as a verbal response, or decision to cooperate).

The core part of the iGEN model will be augmented with a culture-specific component that can respond to culturally-unique features that need to be represented in a training scenario. In addition, it may be necessary to have an actor perform actions in the scenario (such as opening a door in response to a knock, or performing some stand-by activity before the trainee approaches). These would have to be separately encoded in the model for a specific actor.

Interaction Technologies

Training cross-cultural interaction within a VE imposes certain requirements on the interaction techniques used within the VE. A great deal of the interaction, and the opportunities for creating misunderstandings and conflicts, depend on relatively subtle cues, such as physical proximity and posture, gaze direction, use of touch, and tone of voice. Classic examples include:

- "social distance," which in many cultures of interest in this effort are much closer than Western European/American conventions;
- direct gaze (for example, in Hispanic cultures) may be considered inappropriate and evoke mistrust;
- touching objects or people with the left hand is inappropriate in many Arabic and Indian cultures;
- a loud voice may be interpreted differently across cultures as friendliness or inappropriate familiarity.

Unfortunately, detection of many of these cues is problematic given the state of VE technology, and depending on the design of a given VE, may not be feasible in any case. For example, social distance requires both the ability to track locomotion over at least a modest area and the ability to correctly portray distance to the user. In a full Cave system, or an HMD-based system, this may be possible, but in many systems, a single large screen serves as the display and depth cues may be inadequate to present the effect. Gaze and facial expression are difficult to track adequately as well, and are distorted by the presence of HMD's. Tracking limbs, especially hands, in sufficient detail to identify gestures at the level of human interaction is rarely done in VE, because of the amount of instrumentation required. Even when the instrumentation exists, recognition of such gestures is still a significant research problem. Current speech recognition technology is typically designed for computer input, and does not capture, let alone present an interface to obtain loudness, pacing, or intonation.

To further complicate matters, these problems are not all amenable to simple workarounds in a training context (such as using alternate input modalities to provide indirect control). In part, such artificial mechanisms decrease the sense of presence that is key to making the training compelling, and in part, use of such mechanisms creates a greater sense of awareness of the factors being monitored (and it is not clear whether this is beneficial or detrimental to training).

Speech itself provides some problems with respect to multilingual support. Commercial, off-the-shelf recognition engines are provided with capabilities to process a relatively small number of "major" world languages, which may not encompass the set of languages used in training contexts. Moreover, even to the extent that they do, they may not work well with non-native speakers.

Having discussed the difficulties that interaction technology presents, we will outline our approach to overcoming these difficulties. To begin with, the training requirements themselves, for a specific selected region/culture and set of scenarios, must inform the process of selecting interaction technologies for the Phase II effort. Our more in-depth analyses planned for Phase II and involving an anthropologist and cultural expert will identify the firm requirements for interaction technology.

In order to scope the effort appropriately, the plan is to focus the bulk of attention to interaction on speech problems. There are several avenues of approach.

- The speech recognition itself must remain “pluggable” in the system being built, so that the best available software for a particular language set can be substituted for use in a particular training context.
- An English-based version of the training scenario will be maintained for demonstration purposes. It can continue to use commercial speech recognition engines (IBM’s ViaVoice).
- Examine the capabilities of some existing open-source toolkits (such as Sphinx and the Mississippi State University recognizer) to provide speaker-dependent, limited vocabulary speech recognition for use with a foreign-language scenario. The requirement that such recognizers impose to train the recognizer for a subject can be turned to advantage by embedding the training into a language-instruction tutorial.
- Identify whether automated speech recognition and language translation initiatives being undertaken by ONR and DARPA have generated applicable products for this effort.
- Use concatenated, recorded speech for synthesis; this is compatible with the template-based language generation approach and will result in higher quality speech, especially for foreign language.

To the extent that gesture and gaze prove significant to the training, approaches to gesture recognition-based on magnetic tracking, or video imaging and image processing will be investigated. The tradeoffs between the two mechanisms relate to the type of environment (HMD-based, or CAVE/screen-based), locomotion controls and capabilities, and cost.

Development of Fully Immersive Team Training (FITT; see Lampton, et al., 2001) used an HMD-based system coupled with limited tracking of body, head, feet, and one elbow and corresponding hand. This scheme permitted fairly natural locomotion and allowed full rotation of the field-of-view in correspondence with body rotation. This level of instrumentation may fall somewhat short of that required to do effective gesture detection for VECTOR. However, we will evaluate the equipment and implementation costs associated with an augmented version of this for VECTOR Phase II.

An alternative approach using video-based tracking may be suitable for CAVE or screen-based VE's. The video-based approach has certain implications for the type of VR and style of interaction – it implies that the trainee is relatively stationary and faces a fixed direction (as in front of a screen). This means that the type of free rotation permitted in the FITT-style of environment is not possible. It also means that some type of explicit controller (e.g., a handheld joystick, such as the AirStick or CyberStik) is required for locomotion. The video-based approach has advantages in terms of equipment cost, reduced encumbrance of the trainee, and greater flexibility and extensibility for detailed gestures.

Software Architecture

This section describes the software architecture we developed for VECTOR. This discussion provides a high-level view of the software components, including commercially-available components. After presenting the structural view, we will discuss construction details, including platform, integration techniques, and tools.

Components

Figure 2, shows the primary components of VECTOR, at the top level. These components include:

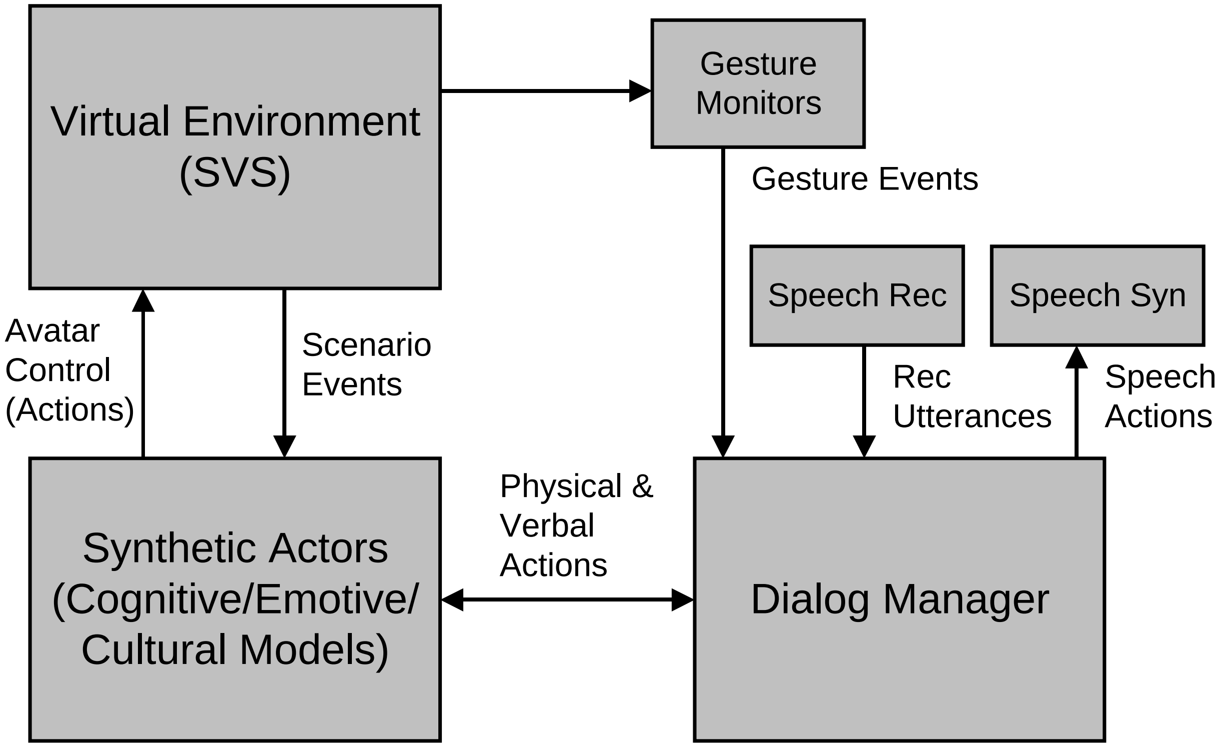

Figure 2. High-level VECTOR architecture.

- Virtual Environment (VE), which will provide the basic VE and support avatar display and movement, as well as supporting tracking and display devices in the immersive context;
- Dialog Manager Component, which will handle details of “face-to-face” verbal interaction with synthetic actors in the virtual environment, as well as the connection with the various body trackers and physical interaction devices;
- Speech Recognition and Synthesis Components, which will incorporate COTS speech recognition and synthesis software and convert spoken language to digital speech messages, and vice versa;
- Synthetic Actor Component, which is where the cultural models are instantiated; this component will perform the cognitive and emotive functions of the human entities in the scenario and drive avatar actions, movements, and speech (the latter via the dialog manager).

There are various input descriptions to these components that define the environment, culture, and specific training scenario. These consist of:

- terrain and object databases to define the VE and avatars,
- the scenario and actor definitions provided to the synthetic actors, which determine the overall cultural attributes of actors instantiated in the scenario as well as providing for time-driven and situational behaviors on the part of the actors, and
- the interaction script, which provides the dialog manager a mapping from raw utterances (coming from the speech recognizer) to semantic content and provides the range of acceptable responses.

The VE and speech processing components are largely COTS, with an adaptation layer provided to interoperate with the other components. However, the Dialog Manager and Synthetic Actor components require more detailed description (in the sections which follow) because they have a fairly complex internal structure.

Dialog Manager

Speech recognition, in its current state, offers highly variable performance, depending on factors ranging from the subject's voice characteristics, to amount of training of the recognizer, to variations based on room noise level and the subject's posture. The assumption needs to be made that some non-trivial percentage of subject utterances will be incorrectly recognized. Furthermore, the dialog in the training contexts we have examined is open-ended and not highly constrained; as well, the subject's level of familiarity with the target language may be limited. These factors serve to further reduce the likelihood that utterances detected will be well-formed.

The dialog manager serves to provide a mapping between the utterances delivered by the speech recognition system and the set of utterances deemed likely for the given training scenario by the interaction script. This is accomplished by use of "fuzzy" matching techniques that select utterances from the interaction script on the basis of a Cartesian product match. The idea is to treat both the "heard" utterance and each utterance of the interaction script as a term vector. The Cartesian (or "dot") product of the term vectors yields the cosine of the "angle" between the vectors. The largest cosine value found (i.e., closest to 1) represents the closest match. This vector space approach has a considerable history of use in information retrieval applications (see Salton, 1989). In addition, it has seen recent application in dialog systems such as AutoTutor (Graesser et al., 1999).

Besides its direct matching of spoken to expected utterances, the dialog manager must incorporate extra-linguistic cues in its processing. Examples of these might be physical gestures (such as pointing, handshakes, touching objects, etc.) and proximity and gaze/body facing cues to identify the intended listener. We plan to represent these as tokens appearing in the interaction script. The gestures themselves are monitored by a separate software component that receives data from the trackers and devices in the VE. The gesture recognizer will potentially need adaptation/extension as future cultural training scenarios are developed that require monitoring of new gestures.

Synthetic Actors and Cultural Models

The synthetic actor component is responsible for representing the behaviors of the simulated human entities. As described above, the synthetic actors represent a hybrid of perceptual, cognitive, and emotive models. The implementation of the perceptual models is through separate perceptual detectors that reside outside of the model, but communicate with the iGEN cognitive model through its perceptual demons. The emotive model should be implemented initially in the iGEN "shell" layer – the software layer that provides an interface to the environment and arbitrary low-level processing. Ultimately, as the emotive model design proves its worth and proper higher-level mechanisms are identified for representing it, the emotive model would be a candidate for inclusion in iGEN proper (as an extension to the modeling language).

As has been described, a capability is desired to be able to adapt generic models for a specific culture, as well as extending them to support "idiosyncratic" cultural features (ones that do not fit within the cultural template, but are still required for training) and instantiating them for individuals. The initial implementation of this will take the form of cultural profiles and scenario scripts. When the scenario script specifies instantiation of a specific actor (whose characteristics are drawn from a specific cultural profile), the model instance is created and reads the profile. The parameters of this profile are adjusted and extended by reading additional parameters specified with the actor definition within the scenario script. These additional parameters represent offsets from the norm represented by the cultural profile, as well as simple extensions that may be used by the cognitive model itself (such as name and gender of the actor). In addition, specialized aspects of the cognitive model may be composed with the main, generic model either by developing a preprocessing step (to extend the generic model with specialized capabilities) or by developing a capability to load additional model code (such as tasks and methods). This introduces the complexity of having to extend the blackboard definition after loading, which may not be feasible, but is worth some investigation. Initially, we will adopt the preprocessing approach.

Integration

The connections depicted in Figure 2 are functional – they depict flows of data between the components. The actual connection mechanism used will be a type of message-oriented middleware. The effect of this is that all of the components are interconnected to one another; the functional linkages depicted in the diagram are created as a by-product of the publish-and-subscribe nature of such middleware. In terms of physical integration, we will use "message-oriented middleware" (such as CORBA and HLA) to connect dialog manager, avatars, and executing model instances. This permits a greater degree of dynamism (instances need not be fixed ahead of time) and permits distribution of computation across machines, if necessary. It also enables a greater degree of flexibility in configuring the system for using different speech recognizers and tracking technology. Our preferred integration mechanism is the DMSO High-Level Architecture (HLA). DIS/HLA gateways have been developed separately by STRICOM and NAWCTSD (SMOC), and can provide support for integrating with DIS-based simulations/environments as well.

While the Phase I work focused on virtual-environment-based implementations of the VECTOR technology, different application areas and training audiences may benefit from other delivery vehicles, including desktop PC-based first-person games and web-delivered content. The Phase I work has defined both an architecture and a development strategy that will decouple the primary VECTOR technology components (the dialog manager and synthetic actors) from the medium of delivery and the interaction mechanisms. This means that it will be feasible to examine the use of these technologies in other contexts, using different interface modalities (such as typing or GUI-based selection).

References

Center for Army Lessons Learned (CALL) (1995). Peace Operations Training Vignettes with Possible Solutions. CALL Newsletter No. 95-2. Fort Leavenworth, KS: U.S. Army Training and Doctrine Command.

CHI Systems, Inc. (2002). *iGEN overview*. Retrieved January 30, 2002, from http://www.cognitiveagent.com/

Collins, A., Brown, J. S., & Newman, S. (1987). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Cognition and Instruction: Issues and agendas*. Hillsdale, NJ: Erlbaum.

Gott, S. P. (1989). Apprenticeship instruction for real-world tasks: The coordination of procedures, mental models, and strategies. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 97-169). Washington, DC: American Educational Research Association.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & TRG (1999). AutoTutor: A simulation of a human tutor. *Journal of Cognitive Systems Research*, 1, 35-51.

Hofstede, G. (1980, 2nd Ed. 1984). *Cultural consequences: International differences in work-related values*. Newbury Park, CA: Sage.

Kreyenbroek, P. & Allison, C. (1996). *Kurdish culture and identity*. Atlantic Highlands, NJ: Zed Books Ltd.

Lampton, D.R., McDonald, D.P., Rodriguez, M.E., Cotton, J.E., Morris, C.S., Parsons, J., and Martin, G. (2001). *Instructional Strategies for Training Teams in Virtual Environments*. (Technical Report 1110, ADA 389674). U.S. Army Research Institute for the Behavioral & Social Sciences, Alexandria, VA.

Meho, L. I. &, Maglaughlin, K. L. (2001). *Kurdish culture and society: An annotated bibliography (Bibliographies and indexes in ethnic studies)*. Westport CT.: Greenwood Publishing Group.

Nash, W. L. (1992). *Operations other than war, Volume 1, Humanitarian assistance*. Newsletter No. 92-6. Fort Leavenworth, KA: The Center for Army Lessons Learned (CALL), U.S. Army Combined Arms Command. Retrieved August 14, 2002, from http://call.army.mil/products/newsltrs/92-6/toc.html

Olsen, H. & Davis, J. (1999). *Training U.S. Army officers for peace operations: Lessons from Bosnia*. (Special Report No. SR-10/99). Washington D.C.: United States Institute of

Peace. Retrieved August 14 2002, from http:// www.usip.org/oc/sr/sr991029 /sr991029.html

Riesbeck, C. K. (1996). Case-based teaching and constructivism: Carpenters and tools. In B.G. Wilson (Ed.), *Constructivist learning environments: Case studies in instructional design*, (pp 49-61). Englewood Cliffs, NJ: Educational Technology Publishing.

Salton, G. (1989). *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Boston, MA: Addison-Wesley.

Velasquez, J. D. (1997). Modeling emotions and other motivations in synthetic agents. *Proceedings of AAAI-97*. pages 10-15. AAAI, AAAI Press and The MIT Press.

Wise, J. C., Hannaman, D. L., Kozumplik, P., Franke, E., & Leaver, B. L. (March, 1997). *Methods to improve cultural communication skills in special operations forces*. Raleigh, NC: North Carolina Center for World Languages & Cultures.

Zachary, W., Le Mentec, J.-C., & Glenn, F. (2000). *Developing computational models of metacognition in support of manning reduction technology development*. Technical Report (000115.9705): Lower Gwynedd, PA: CHI Systems.

Zachary W., Le Mentec, J. C. and Ryder, J., (1996). Interface agents in complex systems. In C. Ntuen & E. Park (Eds.), *Human interaction with complex systems: Conceptual principles and design practice* (pp 35-52). Norwell, MA: Kluwer Academic Publishers.